



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 168272

TO: Konstantina Katcheves
Location: REM/2A60/2C70
Art Unit: 1636
Thursday, February 23, 2006
Case Serial Number: 10/070574

From: Toby Port
Location: Biotech-Chem Library
REM-1A59
Phone: 571-272-2523

toby.port@uspto.gov

Search Notes

Examiner Katcheves,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Toby Port
X22523



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact:*

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☒ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library, Remsen Bldg.



STIC-Biotech/ChemLib

180572

From: Chan, Christina
Sent: Wednesday, February 22, 2006 3:47 PM
To: Katcheves, Konstantina; STIC-Biotech/ChemLib
Subject: RE: After final 10070574

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Katcheves, Konstantina
Sent: Wednesday, February 22, 2006 10:08 AM
To: Chan, Christina
Subject: After final 10070574

Christina,

Would you approve the following search:

Please search nucleotides 574-1134 of SEQ ID NO:1 and nucleotides 28-542¹ of SEQ ID NO:2 against the commercial and interference databases.

Thanks,
Tina

Konstantina Katcheves
Patent Examiner , AU1636
Phone: (571) 272-0768
Room: REM 2A60
Mail: REM 2C70

2/22/2006

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:21:30 ; Search time 4624.57 Seconds
(without alignments)
6317.898 Million cell updates/sec

Title: US-10-070-574-2_COPY_28_541

Perfect score: 514

Sequence: 1 ttattacgctgaagaagat.....aatgatacatgtctctgcag 514

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	541	6	AX099124 Sequence
2	513	99.8	2198	15	KMAJ76
3	127.8	24.9	110000	15	Continuation (9 of
C 4	127.8	24.9	110000	15	Continuation (10 o
5	73.6	14.3	2179	15	AB059425 Kluyverom
C 6	51.8	10.1	196665	9	AC141560 Mus muscu
7	50.8	9.9	349980	6	AX344557 Sequence
C 8	49.4	9.6	591	6	CQ522636 Sequence
C 9	49.4	9.6	3470	6	CQ412028 Sequence
C 10	49.4	9.6	3470	6	CQ492626 Sequence
C 11	49.4	9.6	3470	6	CQ492817 Sequence
C 12	49	9.5	217947	14	AC162562 Bos tauru
13	49	9.5	349980	6	AX344559 Sequence
14	48.8	9.5	16127	6	AX345646 Sequence
15	48.2	9.4	349980	6	AX344566 Sequence
16	48	9.3	349980	6	AX344565 Sequence
17	47.8	9.3	18997	6	AX345472 Sequence
18	47.8	9.3	18997	6	AX347384 Sequence

19	47.8	9.3	18997	6	AX349105 Sequence
20	47.8	9.3	18997	6	AX657772 Sequence
21	47.8	9.3	18997	6	AX659046 Sequence
22	47.8	9.3	141466	9	AC144817 Mus muscu
23	47.8	9.3	207997	9	AC159136 Mus muscu
C 24	47.8	9.3	349751	2	PFMAL4P3 Plasmodi
25	47.6	9.3	1522	2	AY701231 Orconecte
26	47.6	9.3	8392	6	AX346392 Sequence
C 27	47.6	9.3	158574	14	CR932984 Danio rer
C 28	47.6	9.3	159017	5	BX511209 Zebrafish
C 29	47.4	9.2	110000	14	AL954295_4 Continuation (5 of
C 30	47.4	9.2	250823	2	AE014821 Plasmodi
C 31	47.2	9.2	1869	8	BC064898 Homo sapi
32	47.2	9.2	3036	6	CQ807305 Sequence
33	47.2	9.2	6289	6	AX598860 Sequence
34	47.2	9.2	6289	6	AX599006 Sequence
35	47.2	9.2	9289	6	AX251256 Sequence
36	47.2	9.2	9289	6	AX767470 Sequence
37	47.2	9.2	9289	6	AX767546 Sequence
C 38	47.2	9.2	122746	15	CR932958 Medicago
C 39	47.2	9.2	194093	9	AC139037 Mus muscu
C 40	47.2	9.2	198277	14	CR931998 Danio rer
41	47.2	9.2	203384	14	CR932357 Danio rer
42	47	9.1	5397	6	AX347107 Sequence
C 43	47	9.1	215498	14	AC159856 Bos tauru
44	46.8	9.1	7201	6	AX345238 Sequence
45	46.6	9.1	393	6	CQ398211 Sequence

ALIGNMENTS

RESULT 1	AX099124	Sequence 2 from Patent WO0120005.	541 bp	DNA	linear	PAT 02-APR-2001
LOCUS	AX099124					
DEFINITION	AX099124					
ACCESSION	AX099124.1	GI:13538334				
VERSION	AX099124.1	GI:13538334				
KEYWORDS	Kluyveromyces marxianus					
SOURCE	Kluyveromyces marxianus					
ORGANISM	Kluyveromyces marxianus					
REFERENCE	1	Becher, D., Siekstele, R., Bartkeviute, D., Sasnauskas, K., Doehner, L. and Salim, S.				
AUTHORS	Regulatory sequences and expression cassettes for yeasts					
TITLE	Patent: WO 0120005-A 2 22-MAR-2001;					
JOURNAL	TAD Pharmazeutisches Werk GmbH (DE)					
FEATURES	Location/Qualifiers					
source	1..541					
	/organism="Kluyveromyces marxianus"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:4911"					
ORIGIN						
Query Match	100.0%	Score 514;	DB 6;	Length 541;		
Best Local Similarity	100.0%;	Pred. No. 2.6e-96;				
Matches 514;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	TTTATTACGTCGAAGATAGGGAGCTCTCAATGCGGTCTCTGAATGGTTCATTCATT	60			
Db	28	TTTATTACGTCGAAGATAGGGAGCTCTCAATGCGGTCTCTGAATGGTTCATTCATT	87			
QY	61	TCGATACCTCGGGGACTTCCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC	120			
Db	88	TCGATACCTCGGGGACTTCCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC	147			
QY	121	TTTCTTTCTATTGTTTTTGTGTTTTATGGAATATAGCTTTGATGATTAGGATATTTTT	180			
Db	148	TTTCTTTCTATTGTTTTTGTGTTTTATGGAATATAGCTTTGATGATTAGGATATTTTT	207			
QY	181	GTATGTAACCAATACATGCTTCTTAATATATAGTAGAGGTGGGCATCTTCTCTCATTA	240			

Db 208 FTAGTGAACAATAATGCTTGAATAATATAGTACGAGTGGGCAATCTACTCTCATTA 267
Qy 241 TTGGTGTCTTTTATTTAGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 300
Db 268 TTGGTGTCTTTTATTTAGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 327
Qy 301 TATCATGTTCTCTCTTTTGAAGAGGTCCCAATGCTTCCAGATAGCCAGCAATCTTT 360
Db 328 TATCATGTTCTCTCTTTTGAAGAGGTCCCAATGCTTCCAGATAGCCAGCAATCTTT 387
Qy 361 CCATGATATTTTGGCGCTTCTTTGACCTGGTGCACACCTTTTCCGAACCAAGATGCAAGT 420
Db 388 CCATGATATTTTGGCGCTTCTTTGACCTGGTGCACACCTTTTCCGAACCAAGATGCAAGT 447
Qy 421 GCTGCTGATACAAACCTGTATTATCAACAATTTGGATCCATCAGCTCACAATCCACAG 480
Db 448 GCTGCTGATACAAACCTGTATTATCAACAATTTGGATCCATCAGCTCACAATCCACAG 507
Qy 481 CTGAAGATACAGAAAAATGATACATGCTCTCTGCAG 514
Db 508 CTGAAGATACAGAAAAATGATACATGCTCTCTGCAG 541

RESULT 2
KMAJ76 2198 bp DNA linear PLN 15-APR-2005
LOCUS Kluveromyces marxianus epgl gene encoding endopolygalacturonase,
DEFINITION partial.
ACCESSION AJ000076
VERSION AJ000076.1 GI:2597956
KEYWORDS endopolygalacturonase; epgl gene.
SOURCE Kluveromyces marxianus
ORGANISM Kluveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluveromyces.
REFERENCE 1
AUTHORS Siekstele,R., Bartkeviciute,D. and Sasnauskas,K.
TITLE Cloning, targeted disruption and heterologous expression of the
Kluveromyces marxianus endopolygalacturonase gene (EPG1)
JOURNAL Yeast 15 (4), 311-322 (1999)
PUBMED 10206190
REFERENCE 2 (bases 1 to 2198)
AUTHORS Siekstele,R.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1997) Siekstele R., Gene engineering laboratory,
Institute of Biotechnology, V.Graiciuno 8, 2028, LITHUANIAN
REPUBLIC

REMARK revised by author 04-NOV-97
FEATURES
Location/Qualifiers
1..2198
/organism="Kluveromyces marxianus"
/mol_type="genomic DNA"
/strain="BMK Y-719"
/db_xref="taxon:4911"
/chromosome="VI"
338..>1655
/gene="epgl"
338..352
/gene="epgl"
573..>1655
/gene="epgl"
/EC_number="3.2.1.15"
/codon_start=1
/product="endopolygalacturonase"
/protein_id="CAA03900.1"
/db_xref="GI:2597957"
/db_xref="GOA:O13478"
/db_xref="InterPro:IPR000743"
/db_xref="InterPro:IPR006626"
/db_xref="UniProt/TREMBL:O13478"
/translation="MLFSTNLLIAPASALLAEASPLEKRDSCITLSGKTAGGSLNCAT
VTNNVNEPSGKTLDTLGLPDGATVFGQVTFGYDEWVGPLVLSISGRNKKRVGKSGH

LLDGDGARWWDGKGKKPKFMSLRLLTGNSDVGGLOIKNTPIQALISVNSCSDTVIH
DVTIDNRDGDKNLGHNTDPSDYGNNVNTIENSHVYNDQDLAVNSGTGVYKNNYC
SGHGASIGSVGLRSNNVDIVIFENNQIVNSDNLRIKTIQAKATSRVNNVHFLSNTI
SGRKFQIVVETDYSNGSTTGTGSGKVPITNFEDVDLGTGSDVSAYIRKIKIPVAGASKW
TKNVDITGSSSFSGCTGIPSGSAFC"
573..657
/gene="epgl"
658..>1655
/gene="epgl"
/product="endopolygalacturonase"
/EC_number="3.2.1.15"
sig_peptide
mat_peptide
ORIGIN
Query Match 99.8%; Score 513; DB 15; Length 2198;
Best Local Similarity 100.0%; Pred. No. 2.9e-96; Indels 0; Gaps 0;
Matches 513; Conservative 0; Mismatches 0;
Qy 1 TTTTAAACGTGAAGAAGATAAGGAAAGTCTTCAATGCGGTTCTGAATGGTTGATCCATT 60
Db 1686 TTTTAAACGTGAAGAAGATAAGGAAAGTCTTCAATGCGGTTCTGAATGGTTGATCCATT 1745
Qy 61 TCGATACCTCGGGACCTTCTTTGAATATATTCTGAGAGTATGACAGTTGGTTCTTCTTC 120
Db 1746 TCGATACCTCGGGACCTTCTTTGAATATATTCTGAGAGTATGACAGTTGGTTCTTCTTC 1805
Qy 121 TTTCTTTCTATTGTTTGTGTTTATGAAATATAGCTTTTCATGATTTAGGATATTTTTT 180
Db 1806 TTTCTTTCTATTGTTTGTGTTTATGAAATATAGCTTTTCATGATTTAGGATATTTTTT 1865
Qy 181 GTAGTGAACCAATACATGCTTTGATTATATACGTACGAGTGGGCAATTTCTACTCTCATTA 240
Db 1866 GTAGTGAACCAATACATGCTTTGATTATATACGTACGAGTGGGCAATTTCTACTCTCATTA 1925
Qy 241 TTGGTGTCTTTTATGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 300
Db 1926 TTGGTGTCTTTTATGGAGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGGAAACGTAA 1985
Qy 301 TATCCATGTTCTCTCTTTTGAAGAGTCCCAACCAATGCTTCCAGATAGCCAGCAATTTCTT 360
Db 1986 TATCCATGTTCTCTCTTTTGAAGAGTCCCAACCAATGCTTCCAGATAGCCAGCAATTTCTT 2045
Qy 361 CCATGATATTTTGGCGCTTGTGTTGACCTGGTGCACACCTTTTCCGAACCAAGATGCAAGT 420
Db 2046 CCATGATATTTTGGCGCTTGTGTTGACCTGGTGCACACCTTTTCCGAACCAAGATGCAAGT 2105
Qy 421 GCTGCTGATACAAACCTGTATTATCAACAATTTGGATCCATCAGCTCACAATCCACAG 480
Db 2106 GCTGCTGATACAAACCTGTATTATCAACAATTTGGATCCATCAGCTCACAATCCACAG 2165
Qy 481 CTGAAGATACAGAAAAATGATACATGCTCTCTGCA 513
Db 2166 CTGAAGATACAGAAAAATGATACATGCTCTCTGCA 2198

RESULT 3
CR382121.08/c
WPCOMMENT
Sequence split into 11 fragments LOCUS CR382121 Accession CR382121
Fragment Name Begin End
CR382121_00 1 110000
CR382121_01 100001 210000
CR382121_02 200001 310000
CR382121_03 300001 410000
CR382121_04 400001 510000
CR382121_05 500001 610000
CR382121_06 600001 710000
CR382121_07 700001 810000
CR382121_08 800001 910000
CR382121_09 900001 1010000
CR382121_10 1000001 1062590
Continuation (9 of 11) of CR382121 from base 800001 (CR382121 Kluveromyces lactis str:
Query Match 24.9%; Score 127.8; DB 15; Length 110000;

Best Local Similarity 74.9%; Pred. No. 6.5e-17;
Matches 173; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 284 TTAGAGCGGACGTAATATCCATGTTCTTCTTTGAAGAGTCCCAATGCTTCCC 343
Db 105792 TCAAAGTGCAACGTAACATCTTTTCTTGAAGATATCCCAATGTTTATT 105733

QY 344 AGATAGCCAGCATTTCTCCATGATATTTTGGCTTGTGTCACCTGGTGACACCTTTTCG 403
Db 105732 TGATGACCGCATTTCTCCAGATATTTCTAGCCTGTTGACGACTTGATACACCTTTTG 105673

QY 404 AACCAAGATGTCAGTGTGCTGTATACAACTGTTATTATCAATTTCTGATCCAT 463
Db 105672 AATCAAGGAGAGCAGATGTCAGATACAAATGGCTATGATGATTAATTCAGGAGCCAT 105613

QY 464 CAGCTCACAATCCAGCTGAGATACAGAAATGATACATGTTCTTCGAG 514
Db 105612 CAATTCAC-ATCCCAACTGAAATACAAATATATATATATGTTCTCAGCAG 105563

RESULT 4
CR382121_09/C
WPCOMMENT

Sequence split into 11 fragments LOCUS CR382121 Accession CR382121

Fragment Name	Begin	End
CR382121_00	1	110000
CR382121_01	100001	210000
CR382121_02	200001	310000
CR382121_03	300001	410000
CR382121_04	400001	510000
CR382121_05	500001	610000
CR382121_06	600001	710000
CR382121_07	700001	810000
CR382121_08	800001	910000
CR382121_09	900001	1010000
CR382121_10	1000001	1062590

Continuation (10 of 11) of CR382121 from base 900001 (CR382121 Kluyveromyces lactis str)

Query Match 24.9%; Score 127.8; DB 15; Length 110000;
Best Local Similarity 74.9%; Pred. No. 6.5e-17;
Matches 173; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 284 TTAGAGCGGACGTAATATCCATGTTCTTCTTTGAAGAGTCCCAATGCTTCCC 343
Db 5792 TCAAAGTGCAACGTAACATCTTTTCTTGAAGATATCCCAATGTTTATT 5733

QY 344 AGATAGCCAGCATTTCTCCATGATATTTTGGCTTGTGTCACCTGGTGACACCTTTTCG 403
Db 5732 TGATGACCGCATTTCTCCAGATATTTCTAGCCTGTTGACGACTTGATACACCTTTTG 5673

QY 404 AACCAAGATGTCAGTGTGCTGTATACAACTGTTATTATCAATTTCTGATCCAT 463
Db 5672 AATCAAGGAGAGCATGCTGCAATACAAATGGCTATGATGATTAATTCAGGAGCCAT 5613

QY 464 CAGCTCACAATCCAGCTGAGATACAGAAATGATACATGTTCTTCGAG 514
Db 5612 CAATTCAC-ATCCCAACTGAAATACAAATATATATATGTTCTCAGCAG 5563

RESULT 5
AB059425
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AB059425
Kluyveromyces wickerhamii PGW1 gene for polygalacturonase, complete cds.
AB059425.1 GI:15991077
Kluyveromyces wickerhamii
Kluyveromyces wickerhamii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
Kishida, M. and Kawasaki, H.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Cloning of a polygalacturonase gene from Kluyveromyces wickerhamii
Unpublished
2. (bases 1 to 2179)
Kishida, M. and Kawasaki, H.
Direct Submission
Submitted (06-APR-2001) Masao Kishida, Osaka Prefecture University,
Graduate School of Agriculture and Biological Sciences, 1-1
Gakuen-cho, Sakai, Osaka 599-8531, Japan
(E-mail: masayuki@biochem.osakafu-u.ac.jp, Tel: 81-722-54-9455,
Fax: 81-722-54-9921)

FEATURES
source

1. 2179

/organism="Kluyveromyces wickerhamii"

/mol_type="genomic DNA"

/db_xref="taxon:51658"

718. 1803

/gene="PGW1"

718. 1803

/gene="PGW1"

/codon_start=1

/product="polygalacturonase"

/protein_id="BAB69585.1"

/db_xref="GI:15991078"

/translation="MLFNKLIILSSAILTSASPLEKRDSCILTGTGGVNGCSA
VIVQVSVPAKTLDTLGLKSGATVFEGETTSFGQEWELISAGSGITVTCASN
LIDGGRWHDGKNGTKPKLLKLTGDSVSGUKIKNSPRQVSVNOCSDLIH
DITIDNTDGGSGKHNTDGFVGSDDTVTKCKVYNQDDCIATVNSGKNIYQNNYC
SGHGIGISGVSDNVDSVFNENNVDSDNLRITKTDQATGKNNVHFLDNEI
SNIAKYIVVESDYKDGSTTTPGSGVPITNLINGVTGVSASSAKRVKILVAGASSW
KWSNIFKFGQSYSSCEGIPSGSGASC"

ORIGIN

Query Match 14.3%; Score 73.6; DB 15; Length 2179;
Best Local Similarity 58.0%; Pred. No. 2.9e-05;
Matches 130; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 203 ATTAATATAGTACGAGTGGGCAATCTACTCTCATTTATGTTGTTTATTTGAGGAAA 262
Db 1919 ATTAATATATATAAATTTGGGAAATGAATGAATGAATGAATGAATGAATGAAT 1978

QY 263 AATTAATCTAGGAGTATCGTTTAGAGCGCAAGTAATATCCATGTTCTTCTTTGAA 322
Db 1979 AAACATTTTACATGTTTGTGATTAAGCAAGCAATATCTAGATTTTCTCTGAA 2038

QY 323 GAGTCCCAACCATGTTCCAGATAGCAGCATTTCCATGATATTTTCGCTTTGTTT 382
Db 2039 TATATCCCAACCATGTTTCTTCTTAACAGCATTTCTCCATTTCTCTTTGCTTTT 2098

QY 383 TGCACGTGTGACACCTTTTCGACCAAGATGTCAGTGTGCT 426
Db 2099 AGCATTGAAACACCTTCCGAATTAGAGATTTAAATGACTCT 2142

RESULT 6
AC141560/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC141560
Mus musculus BAC clone RP23-425A5 from 12, complete sequence.
AC141560
AC141560.4 GI:51854645
HTG.
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 196665)
Nguyen, C. and Haakenson, W.
The sequence of Mus musculus BAC clone RP23-425A5
Unpublished (2001)
2 (bases 1 to 196665)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (17-MAR-2003) Genome Sequencing Center, 4444 Forest Park

```

REFERENCE
AUTHORS      Parkway, St. Louis, MO 63108, USA
TITLE        3 (bases 1 to 196665)
JOURNAL      Wilson.R.K.
SUBMITTED    Submitted (16-JUN-2004) Genome Sequencing Center, 4444 Forest Park
JOURNAL      Parkway, St. Louis, MO 63108, USA
REFERENCE    4 (bases 1 to 196665)
AUTHORS      Wilson.R.K.
TITLE        Direct Submission
JOURNAL      Submitted (02-SEP-2004) Genome Sequencing Center, 4444 Forest Park
REFERENCE    5 (bases 1 to 196665)
AUTHORS      Wilson.R.K.
TITLE        Direct Submission
JOURNAL      Submitted (27-JAN-2005) Genome Sequencing Center, Washington
SUBMITTED    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
COMMENT      On Sep 2, 2004 this sequence version replaced gi:48762631.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: M_BA0425A05
-----

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC library has been constructed by Kazutoyo Osegawa
and Minako Tatenio in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
FEATURES
source
    1..196665
        Location/Qualifiers
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="12"
            /map="12"
            /clone="RP23-425A5"
            /clone_lib="RPCI-23"
repeat_region
    1..55
        /rpt_family="Alu"
repeat_region
    68..248
        /rpt_family="B2"
repeat_region
    299..668
        /rpt_family="L1"
repeat_region
    674..879
        /rpt_family="L1"
repeat_region
    994..1094
        /rpt_family="B2"
repeat_region
    1116..1234
        /rpt_family="L1"
repeat_region
    1574..1642
        /rpt_family="B4"
repeat_region
    2798..2936
        /rpt_family="Alu"
repeat_region
    3148..3215
        /rpt_family="ID"
repeat_region
    3216..3229
        /rpt_family="B4"
repeat_region
    3825..3955
        /rpt_family="MER1_type"
repeat_region
    4010..4155
        /rpt_family="Alu"
repeat_region
    4275..4422
        /rpt_family="Alu"
repeat_region
    5113..5173
        /rpt_family="MER1_type"
repeat_region
    5653..5768
        /rpt_family="Alu"
repeat_region
    5999..6141
        /rpt_family="Alu"
repeat_region
    6151..6343
        /rpt_family="B2"
repeat_region
    6494..6618
        /rpt_family="Alu"
repeat_region
    6619..6754
        /rpt_family="B4"
repeat_region
    6786..6800
        /rpt_family="B4"
repeat_region
    6801..6864
        /rpt_family="ID"
repeat_region
    6993..7182
        /rpt_family="Alu"
repeat_region
    7627..7849
        /rpt_family="MER1_type"
repeat_region
    8112..8201
        /rpt_family="ERV1"
repeat_region
    8344..8699
        /rpt_family="MaLR"
repeat_region
    8952..9007
        /rpt_family="ERV1"
repeat_region
    9724..9760
        /rpt_family="MaLR"
repeat_region
    9761..9881
        /rpt_family="ERV1"
repeat_region
    9893..10224
        /rpt_family="MaLR"
repeat_region
    10556..10694
        /rpt_family="B2"
repeat_region
    11080..11282
        /rpt_family="B2"
repeat_region
    11324..11495
        /rpt_family="B2"
repeat_region
    11518..11616
        /rpt_family="B4"
repeat_region
    11985..12519
        /rpt_family="MaLR"
repeat_region
    12768..12832
        /rpt_family="L1"
repeat_region
    13018..13137
        /rpt_family="L1"
repeat_region
    13163..13218
        /rpt_family="Alu"
repeat_region
    16855..17063
        /rpt_family="B2"
repeat_region
    17192..17455
        /rpt_family="B4"
repeat_region
    18530..18636
        /rpt_family="Alu"
repeat_region
    18732..18803
        /rpt_family="B4"
repeat_region
    19179..19988
        /rpt_family="L1"

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:20:52 ; Search time 450.886 Seconds
(without alignments)
7597.608 Million cell updates/sec

Title: US-10-070-574-2_COPY_28_541

Perfect score: 514

Sequence: 1 ttattacgtgaagaagat.....aatgatactgtctctgcag 514

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	541	5	Aaf61990 K. marxia
C 2	49.4	9.6	591	5	ABV54484 Human pro
C 3	49.4	9.6	3470	5	ABV24695 Human pro
C 4	49.4	9.6	3470	5	ABV24504 Human pro
C 5	49.4	9.6	3470	5	ADL45209 Human ova
6	48.8	9.5	16127	6	ABL32744 Human imm
7	47.8	9.3	18997	6	ABL32570 Human imm
8	47.8	9.3	18997	6	ABK33948 Human DNA
9	47.8	9.3	18997	8	ADA20352 Prostate
10	47.8	9.3	18997	8	ADA84159 Human ren
11	47.6	9.3	8392	6	ABL33490 Human imm
C 12	47.2	9.2	2000	11	ACL37108 Rice stre
13	47.2	9.2	3036	13	ADS9739 Oligonuc
14	47.2	9.2	6289	8	ABZ10206 Haematopo
15	47.2	9.2	6289	8	ABZ10060 Haematopo
16	47.2	9.2	9289	4	AAS46502 Tumour su
17	47.2	9.2	9289	10	ADE84198 Human lym
18	47.2	9.2	9289	10	ADE84122 Human lym
19	47	9.1	5397	6	ABL34205 Human imm

20	46.8	9.1	7201	6	ABL32336	Human imm
21	46.6	9.1	393	5	ADI72540	Human ova
22	46.6	9.1	393	5	ADL37679	Human ova
23	46.6	9.1	497	5	ADL37664	Human ova
24	46.6	9.1	497	5	ADI72525	Human ova
25	46.6	9.1	5474	6	ABL33370	Human imm
26	46.6	9.1	8147	6	ABL32428	Human imm
27	46.6	9.1	113515	6	ABL34174	Human imm
28	46.2	9.0	6316	6	ABL33639	Human imm
29	46	8.9	6182	6	ABL49388	Human imm
C 30	45.6	8.9	15649	6	ABL70544	Human pol
C 31	45.6	8.9	5949	14	ADY18488	DNA encod
C 32	45.4	8.8	2000	11	ACL35887	Chemical
33	45.4	8.8	5771	6	ABN80073	Rice stre
34	45.4	8.8	9905	6	ABL32062	Human che
C 35	45.2	8.8	2355	12	ADP18734	Human imm
36	45.2	8.8	5034	13	ADS89648	Mouse lip
37	45.2	8.8	6577	4	AAS46718	Oligonuc
38	45.2	8.8	10286	4	AAS45308	Tumour su
39	45.2	8.8	10286	6	ABK28147	Chemical
C 40	45	8.8	526	6	ABQ39217	DNA trans
41	45	8.8	526	6	ABQ39216	Oligonuc
C 42	45	8.8	1221	11	ADL33377	Oligonuc
43	45	8.8	3326	13	ADS89718	Human tra
44	45	8.8	3326	13	ADS89444	Oligonuc
45	45	8.8	7316	6	ABN80240	Human che

ALIGNMENTS

RESULT 1

AAF61990 ID AAF61990 standard; DNA; 541 BP.

XX AAF61990;

XX 22-AUG-2001 (first entry)

XX K. marxianus endopolygalacturonase terminator region DNA SEQ ID 2.

XX Endopolygalacturonase; promoter; hepatitis B surface antigen; pectin;

XX polyoma virus VP1; staphylococcal protein A; vaccine; antiviral;

XX antibacterial; regulatory region; terminator; ds.

XX Kluyveromyces marxianus.

XX WO200120005-A1.

XX 22-MAR-2001.

XX 05-SEP-2000; 2000WO-EP008662.

XX 10-SEP-1999; 99DE-01043383.

XX (TADP-) TAD PHARMA GMBH.

XX Becher D, Siekstele R, Bartkeviciute D, Sasnauskas K, Doehner L;

XX Salim S;

XX WPI; 2001-244812/25.

XX New promoter from Kluyveromyces marxianus, useful for controlling expression of foreign genes in yeast, e.g. production of vaccine antigens.

XX Claim 2; Page 27; 32pp; German.

XX This invention describes a novel promoter sequence (S1), isolated from Kluyveromyces marxianus. (S1) is a promoter for controlling expression of foreign proteins in yeast, particularly hepatitis B surface antigen, CC polyoma virus VP1 or staphylococcal protein A for use in vaccines. (S1) is a strong promoter that can be induced by adding pectin to the culture

CC medium and when combined with its native signal sequence can provide
CC secretion of heterologous protein. Cassettes based on (S1) are stably
CC retained in the chromosome when cells are grown under optimal conditions
CC and provide high yields of recombinant protein. Kluyveromyces marxianus
CC is generally regarded as safe, can use a wide variety of carbon sources,
CC is not very sensitive to temperature (it can grow at up to 45 plusOC) and
CC under ideal conditions has a doubling time of only 35 minutes. The
CC products of the invention have antiviral and antibacterial activity. This
CC sequence represents a terminator region of the Kluyveromyces marxianus
CC endopolygalacturonase described in the disclosure of the invention
XX

SQ Sequence 541 BP; 138 A; 96 C; 104 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 514; DB 5; Length 541;

Best Local Similarity 100.0%; Pred. No. 3.1e-114;

Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTTATTAACTGAAGAATAAGGAAGTCTTCAATCGGTTCTGAATGGTTGATCCATT 60

Db TTTATTAACTGAAGAATAAGGAAGTCTTCAATCGGTTCTGAATGGTTGATCCATT 87

Qy 61 TCGATACCTCGGGACCTCTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC 120

Db TCGATACCTCGGGACCTCTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTTC 147

Qy 121 TTTCTTTCTATGTTTGTGTTTATGGAATATAGCTTTGATCATTTAGGATATTTTT 180

Db TTTCTTTCTATGTTTGTGTTTATGGAATATAGCTTTGATCATTTAGGATATTTTT 207

Qy 181 GTAGTGAACCAATACATGCTTGATTAATATACGTACGAGTGGGCATTCTACTCTCATTA 240

Db GTAGTGAACCAATACATGCTTGATTAATATACGTACGAGTGGGCATTCTACTCTCATTA 267

Qy 241 TTGGTGTTTATTGGAGGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGCGAAGTAA 300

Db TTGGTGTTTATTGGAGGGAAAAATTAATCTAGGAGTATCGTTTAGAGCGCGAAGTAA 327

Qy 301 TATCCATGTTCTCTCTTGAAGAGTCCACCAATCTCTCCAGATAGCCAGCATTCCT 360

Db TATCCATGTTCTCTCTTGAAGAGTCCACCAATCTCTCCAGATAGCCAGCATTCCT 387

Qy 361 CCATGATATTTGGCGTGTGTTTGGCACTGGTGACACCCCTTCGAACCAAGATGTCAAGT 420

Db CCATGATATTTGGCGTGTGTTTGGCACTGGTGACACCCCTTCGAACCAAGATGTCAAGT 447

Qy 421 GCTGCTGATACAACTGTAATTCATACAAATTCGATCATCAGCTCAATCCACAG 480

Db GCTGCTGATACAACTGTAATTCATACAAATTCGATCATCAGCTCAATCCACAG 507

Qy 481 CTGAAGATACAGAAATGATACATGTCTCTGCAG 514

Db CTGAAGATACAGAAATGATACATGTCTCTGCAG 541

RESULT 2

ABV54484/c

ID ABV54484 standard; cDNA; 591 BP.

XX

AC ABV54484;

XX

DT 17-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 54475.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA

XX Schlegel R, Endege WO, Monahan JB;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

PT for detecting presence of prostate cancer, stage of prostate cancer.

XX

PS Claim 1; Page 10531; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification or its complement. (f) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)

CC assessing the aggressiveness or indolence of prostate cancer in a patient

CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

SQ Sequence 591 BP; 230 A; 129 C; 120 G; 112 T; 0 U; 0 Other;

Query Match 9.6%; Score 49.4; DB 5; Length 591;

Best Local Similarity 58.5%; Pred. No. 0.041;

Matches 86; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 77 TTCCTTTGAATATATCTGAGAGTATGACAGTTGGTTTCTTCTTCTTCTTCTTCTT 136

Db TTTTCTTAATT 244

Qy 137 TTGTTTTTATGGAATATAGCTTTGATGATTTAGGATTTTTTTGTAGTGAACCAATACA 196

Db TTTTCTTAATT 244

Qy 197 TGCTTGATTAATATACGTACGAGGTGG 223

Db TGCTGATTTATTACAGGGATAAGATGG 157

RESULT 3

ABV24695/c

ID ABV24695 standard; cDNA; 3470 BP.

XX

AC ABV24695;

XX

DT 16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 24686.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US005171.

Db 5974 AGTAATTGAGGATTTAGAAGGTATTATGTTTGTATTATTTTGTGTTGTTGGAT 6033

Qy 263 AATTAAAT 270

Db 6034 ATGTAAT 6041

Search completed: February 23, 2006, 01:52:10
Job time : 453.886 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:40:58 : Search time 3102.17 Seconds
(without alignments)
7752.175 Million cell updates/sec

Title: US-10-070-574-2_COPY_28_541

Perfect score: 514

Sequence: 1 ttattacacggaagaat.....aatgatacatgtctctcag 514

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_est3.*

4: gb_hcc.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_est7.*

9: gb_gss1.*

10: gb_gss2.*

11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.2	17.0	515	9 AZ928077	AZ928077 479.dif04
2	77.8	15.1	986	11 CNS06N3D	AL406271 T3 end of
3	53.8	10.5	779	11 CNS04A9H	AL281582 Tetradon
4	53.6	10.4	950	11 CNS07185	AL424587 T3 end of
5	52	10.1	975	9 BH179465	BH179465 014_P_10-
6	49.8	9.7	767	10 CNS00A0X	AL055924 Drosophil
7	49.6	9.6	1051	8 DN569917	DN569917 94034428
8	49.4	9.6	1175	2 BE421167	BE421167 HWM006.E0
9	48.8	9.5	325	10 CNS0099P	AL053301 Drosophil
10	48.8	9.5	1094	10 CNS012FZ	AL101513 Drosophil
11	48.8	9.5	1101	10 CNS001GJ	AL050877 Drosophil
12	48.4	9.4	452	2 B1142777	B1142777 rK73g10.Y
13	48.4	9.4	1083	9 BZ696101	BZ696101 SP_Ba007
14	48.4	9.4	1201	10 CNS0167M	AL106396 Drosophil
15	48	9.3	475	9 BH843983	BH843983 TC3-55H10
16	48	9.3	673	9 BH842470	BH842470 TC3-55H9
17	47.8	9.3	759	10 CNS007BZ	AL066974 Drosophil
18	47.6	9.3	284	3 BM882253	BM882253 r27e12.Y
19	47.6	9.3	348	1 A1155694	A1155694 ue03a05.I
20	47.6	9.3	652	7 CO070912	CO070912 GR_Ba28D
21	47.6	9.3	822	6 CB849204	CB849204 MSA-0892
22	47.6	9.3	869	10 CNS017Z2	AL108680 Drosophil

C 23	47.6	9.3	919	10 CNS005RL	AL061409 Drosophil
C 24	47.6	9.3	1334	5 BU849682	BU849682 AGENCOURT
C 25	47.4	9.2	376	3 BM037194	BM037194 fu84c02.Y
C 26	47.4	9.2	724	9 BZ086084	BZ086084 l1e26a01.
C 27	47.4	9.2	801	10 AG034099	AG034099 Pan tTogl
C 28	47.4	9.2	946	3 BI911499	BI911499 603064008
C 29	47.4	9.2	1101	10 CNS017SO	AL108450 Drosophil
C 30	47.2	9.2	950	8 DN584697	DN584697 92277725
C 31	47.2	9.2	1375	10 AG350207	AG350207 Mus muscu
C 32	47	9.1	680	9 BH179930	BH179930 015_L_18-
C 33	47	9.1	680	11 CNS07LWI	AL161684 T3_end of
C 34	46.8	9.1	363	5 BU761068	BU761068 sage2e07.
C 35	46.6	9.1	295	5 BU529209	BU529209 AGENCOURT
C 36	46.6	9.1	929	8 DN584074	DN584074 93954091
C 37	46.6	9.1	994	10 CNS0145S	AL103738 Drosophil
C 38	46.6	9.1	1101	10 CNS0039G	AL063921 Drosophil
C 39	46.6	9.1	1127	11 CNS06FTQ	AL396456 T7 end of
C 40	46.6	9.1	1200	10 CL031866	CL031866 CH216-33I
C 41	46.4	9.0	786	10 CNS00609	AL065615 Drosophil
C 42	46.4	9.0	902	10 CNS010WI	AL099516 Drosophil
C 43	46.4	9.0	1089	10 CL511283	CL511283 SAIL_844
C 44	46.4	9.0	1101	10 CNS002C3	AL097485 Drosophil
C 45	46.2	9.0	1101	10 CNS010D3	AL098817 Drosophil

ALIGNMENTS

RESULT 1
AZ928077 515 bp DNA linear GSS 01-APR-2001
LOCUS 479.dif04c06.s1 Saccharomyces kluyveri genomic clone 479.dif04c06.s1, genomic survey sequence.
DEFINITION
ACCESSION AZ928077.1 GI:13498982
VERSION 1
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri
ORGANISM Saccharomyces kluyveri
REFERENCE 1 (bases 1 to 515)
AUTHORS Cliften,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish,W.R., Waterston,R.H. and Johnston,M.
TITLE Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis
JOURNAL Unpublished (2001)
COMMENT Contact: Johnston M
Department of Genetics
Washington University Medical School
Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA
Tel: 314 362 2735
Fax: 314 362 7855
Email: mj@genetics.wustl.edu
Class: random plasmid subclone.
FEATURES
source
1..515
/organism="Saccharomyces kluyveri"
/mol_type="genomic DNA"
/strain="NRRL Y-12651 (CBS 3082)"
/db_xref="taxon:4934"
/clone="479.dif04c06.s1"
/clone_lib="Saccharomyces kluyveri"
/note="Random genomic sequence"

ORIGIN
Query Match 17.0%; Score 87.2; DB 9; Length 515;
Best Local Similarity 63.5%; Pred. No. 2.5e-10;
Matches 148; Conservative 0; Mismatches 84; Indels 1; Gaps 1;
QY 282 GTTTAGAGCGCAACGTATATCCATGTTCTTCTTTTGAAGAGTCCACCATGCTTC 341
|||||
DB 281 GTTTAAACTGCAACGTTATGTCGAGTTTCTCTCCCTCAAAATGTCCTCCACCAAGCTTT 340
|||||

QY 342 CCAGATAGCAGCATCTTCCATGATATTTGGCTTGTGTTGCACTGTGACACCCCTTT 401
 DB 341 TTCTGACTCCACATTTCTGCATGATGTTTTCCTGTGAGNACTACTCATACCACT 400
 QY 402 CGAACCAAGATGTCAGTCTGCTGCTGATACAAACCTGTATTCATACAAATTCCTGATCC 461
 DB 401 CTATTCAAAGCATCGATATGTGTCATATACACTTCACGACATTTCTTTCAACGATTCATCC 460
 QY 462 ATCAGCTCACATCCACAGCTGAGATACAGAAATGATACATGCTCTGCGAG 514
 DB 461 ATTATCTCAC-AGCCACAGCAGAAATAACAGAAAAGGATGCAACATTTCTGCGAG 512

RESULT 2
 CNS06N3D 986 bp DNA linear GSS 17-JUN-2001
 T3 end of clone AU0AA009B10 of library AU0AA from strain CBS 3082
 of Saccharomyces kluyveri, genomic survey sequence.

ACCESSION AL406271.1 GI:12170067
 VERSION AL406271.1
 KEYWORDS Saccharomyces kluyveri
 SOURCE Saccharomyces kluyveri
 ORGANISM Saccharomycetales; Ascomycota; Saccharomycotina; Saccharomycetes;
 Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
 1 (bases 1 to 986)
 Souci  t,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.

REFERENCE 1
 Authors Wincker,P. and Weissenbach,J.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 986)
 Authors Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
 Gaillardin,C. and Casaregola,S.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 9.

JOURNAL Saccharomyces kluyveri
 PUBMED 11152884
 REFERENCE 3 (bases 1 to 986)
 Authors Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source Location/Qualifiers
 1. .986
 /organism="Saccharomyces kluyveri"
 /mol_type="genomic DNA"
 /strain="CBS 3082"
 /db_xref="taxon:4934"
 /clone="AU0AA009B10"
 /clone_lib="AU0AA"
 /note="end : T3"

ORIGIN

Query Match 15.1%; Score 77.8; DB 11; Length 986;
 Best Local Similarity 63.9%; Pred. No. 4.9e-08;
 Matches 149; Conservative 0; Mismatches 82; Indels 2; Gaps 2;

QY 282 GTTTAGCGCGAAGCTAATATCCATGTTCTTCTTTGAAGAGTCCACCATTGCTTC 341
 DB 229 GTTTAACTGCAACGTTATGTCCAGTTTTCCTCTCAAAATGTCACCAAGCTTT 288
 QY 342 CCAGATAGCCAGCATTTTCCATGATATTTTGGCGTCTTTTGGCATCGTGACACCCCTTT 401
 DB 289 TTCTGACTCCAAACATTCCTCGCATGATGTTTGGCTG- TAAGCACTACTCATACCACT 347
 QY 402 CGAACCAAGATGTCAAGTCTGCTGTATACAAACCTGTATTTCATACAAATTCCTGATCC 461
 DB 348 CTATTCAAAGCATCGATATGTGCCATATACACTTCCACGACATTTCTTTCAACGATTCATCC 407
 QY 462 ATCAGCTCACATCCACAGCTGAGATACAGAAATGATACATGCTCTGCGAG 514
 DB 408 ATTATCTCAC-AGCCACAGCAGATTAACAGAAAAGGATGCAACATTTCTGCGAG 459

RESULT 3

CNS04A9H/c
 LOCUS
 DEFINITION

ACCESSION AL281582
 VERSION AL281582.1 GI:8019905
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis
 ORGANISM Tetraodon nigroviridis

REFERENCE 1
 Authors Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
 PUBMED 10835645
 REFERENCE 2 (bases 1 to 779)
 Authors Genoscope.

TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large
 scale clone-and sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source Location/Qualifiers
 1. .779
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="095G17"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0BG095AD09SP1
 end : PUC-Ori"

ORIGIN

Query Match 10.5%; Score 53.8; DB 11; Length 779;
 Best Local Similarity 35.0%; Pred. No. 0.029;
 Matches 118; Conservative 66; Mismatches 148; Indels 5; Gaps 1;

```

QY 48 TGGTTGATCCATTTCGATACCTCGGGGACTTCCTTTGAATATATTCTGAGAGTATGACAG 107
Db 339 TTGKTTTTCCTTCTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 280
QY 108 TTGGTTTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 167
Db 279 TTTATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 220
QY 168 TAGGATATTTTGTAGTGAACCAATACATGCTTGAATTAATATACGTACGAGGTGGGAT 227
Db 219 TACTGTTTTTATTTATGCGKCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 160
QY 228 TCTACTCTCATTTATTTGTTTATTTGAGGGGAAAAATTAATATCTAGGAGTATGTTTATG 287
Db 159 KKTITWGCATTTTWTGACATTTTWTATGCTTTATTKAT-----KWAGTTACTTTTWTG 105
QY 288 AGCGGAAACGTAATATCATGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCCAGAT 347
Db 104 TKGACAKTKTTTTATTTKTTKTTTTTTTTTTTTCACACCKTTCATKTTTTCATKCTTKAT 45
QY 348 AGCAGCATTTCTCATGATATTTTGGCTGTTTGTG 384
Db 44 CCACATCAKTTTCCCGGGGATGTTGCTGCTGTATAG 8

```

```

RESULT 4
CNS07185/5
LOCUS 950 bp DNA linear GSS 06-JUL-2001
DEFINITION T3 end of clone XAZ0AA001C07 of library XAZ0AA from strain CBS 712
of Kluyveromyces marxianus, genomic survey sequence.
ACCESSION AL424587
VERSION AL424587.1 GI:12207781
KEYWORDS GSS.
SOURCE Kluyveromyces marxianus
ORGANISM Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
1 (bases 1 to 950)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
11152876
2 (bases 1 to 950)
Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
and Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 12.
Kluyveromyces marxianus var. marxianus
FEBS Lett. 487 (1), 71-75 (2000)
11152887
3 (bases 1 to 950)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..950
/organism="Kluyveromyces marxianus"
/mol_type="genomic DNA"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="XAZ0AA001C07"
/clone_lib="XAZ0AA"
/note="end : T3"

```

```

RESULT 5
BH179465/c
LOCUS 975 bp DNA linear GSS 29-OCT-2001
DEFINITION 014 P10-21 SmbAC1 Schistosoma mansoni genomic clone 014P10 5',
genomic survey sequence.
ACCESSION BH179465
VERSION BH179465.1 GI:16280351
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeidida; Schistosomatidae; Schistosoma.
1 (bases 1 to 975)
Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
Williams,D.L., Johnston,D., Lloverde,P.T. and Le Paslier,D.
Construction and characterization of a Schistosoma mansoni
bacterial artificial chromosome library
Genomics 65 (2), 87-94 (2000)
10783255
Other_GSSs: 014_P10-rev
Contact: Pierce RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA014DH05CP1 Bases 255-407 have 95% identity
to S.mansoni EST AA269074.1 from base 1-153. Bases 30-538 have 92%
identity to S.mansoni EST BF936764.1 from base 512-6.
Plate: 014 row: P column: 10
Seq primer: M13 -21 primer
High quality sequence stop: 975.
Location/Qualifiers
1..975
/organism="Schistosoma mansoni"

```

```

FEATURES
source
1..975
/organism="Schistosoma mansoni"

```


[illegible]

RESULT 13
BZ696101
LOCUS
DEFINITION
SP_Ba0073006.f.SP_Ba Sorghum proplequinum genomic clone
SP_Ba0073006 5', genomic survey sequence.
1083 bp DNA linear GSS 02-JUL-2000
BZ696101
ACCESSION
BZ696101
VERSION
BZ696101.1 GI:28389813
KEYWORDS
GSS.

```

FORWARD: att agc ggc cgc gat cc
BACKWARD: gta aaa cga cgc cca gtg
Plate: 0073 row: 0 column: 06
Seq primer: atc agc ggc cgc gat cc
Class: BAC ends.

```

```

/mol_type="genomic DNA"
/db_xref="taxon:132711"
/clone="SP_Ba0073006"
/clone_lib="Sp_Ba"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Paterson lab BAC library (HindIII)"

```

697	TTTTTTTTTTTTTTAATAGGAAGAAATATTTTAAATATATATGTAATATATTTTTATTTTT	Db
121	TTTCTTCTATTTGTTTTTTTATGGAATATAGCTTTGATGATTAGGATATTTTTT	QY
757	GTTTTNNNTAGTTAAATAGTTTATGTGGGANTATTTTTTATNTTAAINTTTTTTTTT	Db
181	GTATGGAACCAATACATGCTTGATTAAATATAGTAGGAGGTGGGCATTTCTACTCTCATTA	QY


```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on:      February 23, 2006, 01:21:30 ; Search time 5047.43 Seconds
              (without alignments)
              6317.898 Million cell updates/sec

Title:      US-10-070-574-1_COPY_574_1134
Perfect score: 561
Sequence:    1 agacgcgtttgcttacag.....cacagattaagctcagaac 561

Scoring table:  IDENTITY NUC
                  Gapop 10.0 , Gapext 1.0

Searched:      5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	561	100.0	1146	6	AX099123	AX099123 Sequence
2	560	99.8	2198	15	KNAJ76	AF000076 Kluverom
3	101.8	18.1	110000	2	AC116957_1	Continuation (2 of
4	96.6	17.2	333321	2	AC116986	AC116986 Dictyoste
5	96.2	17.1	182870	2	AC116960	AC116960 Dictyoste
C 6	96	17.1	110000	2	AC116984	AC116984 Dictyoste
C 7	95.6	17.0	190196	5	CR293524_0	CR293524 Zebrafish
C 8	94.6	16.9	4865	2	AY232273	AY232273 Dictyoste
C 9	93.8	16.7	36148	2	AC116310	AC116310 Dictyoste
10	93.6	16.7	132254	2	AC116330	AC116330 Dictyoste
11	93.6	16.7	214957	14	CR626875	CR626875 Danio rer
12	92.6	16.5	110000	2	AC116305_2	Continuation (3 of
13	92.6	16.5	125623	2	AC115599	AC115599 Dictyoste
14	92.4	16.5	148570	2	AC117076	AC117076 Dictyoste
C 15	92	16.4	2400	2	AF036170	AF036170 Dictyoste
C 16	91.8	16.4	30849	2	AC117082	AC117082 Dictyoste
17	91.8	16.4	25769	2	AC117267	AC117267 Dictyoste
C 18	91.8	16.4	268544	2	AC116956	AC116956 Dictyoste

19	91.6	16.3	144400	15	AY710429	AY710429 Cryptococ
20	91.6	16.3	183648	2	AC117081	AC117081 Dictyoste
c 21	91.4	16.3	7539	2	AY159036	AY159036 Dictyoste
c 22	91.4	16.3	39984	2	AC114258	AC114258 Dictyoste
23	91.4	16.3	110000	2	AC116305	AC116305 Dictyoste
c 24	91.4	16.3	110000	2	AC116305 ⁰	AC116305 Dictyoste
c 25	91.4	16.3	154071	2	AC115598 ⁻	AC115598 Dictyoste
c 26	91.4	16.3	331039	2	AC116979	AC116979 Dictyoste
27	91.2	16.3	317	15	AY022605	AY022605 Dictyoste
c 28	91.2	16.3	82139	2	AC115684	AC115684 Dictyoste
c 29	91.2	16.3	110000	15	AP008213_072	Continuation (73 o
c 30	91.2	16.3	144076	15	AP005754	AP005754 Oryza sat
c 31	91.2	16.3	154091	14	AP003819	AP003819 Oryza sat
c 32	91	16.2	33651	2	AC114261	AC114261 Dictyoste
c 33	91	16.2	179015	5	CR293530	CR293530 Zebrafish
c 34	90.8	16.2	61052	14	AC123513	AC123513 Dictyoste
c 35	90.8	16.2	110000	15	AP008211_115	Continuation (116
c 36	90.8	16.2	135323	15	AC136229 ⁻	AC136229 Oryza sat
c 37	90.8	16.2	136240	2	AC117070	AC117070 Dictyoste
c 38	90.8	16.2	257109	2	AC116963	AC116963 Dictyoste
c 39	90.6	16.1	110000	2	AC116305 ³	Continuation (4 of
c 40	90.6	16.1	229517	14	BX957338	BX957338 Danio rer
c 41	90.4	16.1	254733	2	AC117075	AC117075 Dictyoste
c 42	90.2	16.1	2108	2	AF272790	AF272790 Hirudo ni
c 43	90.2	16.1	105682	2	AC116957 ³	Continuation (4 of
c 44	90.2	16.1	110000	14	BX980561 ⁰	BX980561 Danio rer
c 45	90.2	16.1	110000	15	AP008207_102	Continuation (103
ALIGNMENTS						
RESULT 1	AX099123	AX099123	Sequence 1 from Patent WO0120005.	1146 bp	DNA	linear PAT 02-APR-2001
LOCUS	AX099123	AX099123				
DEFINITION	AX099123	AX099123				
ACCESSION	AX099123	AX099123				
VERSION	AX099123.1	GI:13538333				
KEYWORDS						
SOURCE						
ORGANISM						
	Kluyveromyces marxianus					
	Kluyveromyces marxianus					
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.					
REFERENCE						
AUTHORS	Becher, D., Siekstele, R., Battkeviolute, D., Sasnauskas, K., Doehner, L., and Salim, S.					
TITLE	Regulatory sequences and expression cassettes for yeasts					
JOURNAL	Patent: WO 0120005-A 1 22-MAR-2001;					
TAD	Pharmazeutisches Werk GmbH (DE)					
FEATURES						
source	1. .1146					
	/organism="Kluyveromyces marxianus"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:4911"					
ORIGIN						
Query Match	100.0%;	Score 561;	DB 6;	Length 1146;		
Best Local Similarity	100.0%;	Pred. No. 6.4e-75;				
Matches 561;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps						
QY	1	AGCAGCGGTTTGCTTACAGTTCTTCATGCAATCAGCCAGATTTTGCAATAGTATTAAAC	60			
Db	574	AGCAGCGGTTTGCTTACAGTTCTTCATGCAATCAGCCAGATTTTGCAATAGTATTAAAC	633			
QY	61	TTAGAAATTAAGGCAACATCTTTGGATATGCAATGATAGTACGTTTCGAAACCATATTATT	120			
Db	634	TTAGAAATTAAGGCAACATCTTTGGATATGCAATGATAGTACGTTTCGAAACCATATTATT	693			
QY	121	ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAGTATTATTGAAATTGTT	180			
Db	694	ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAGTATTATTGAAATTGTT	753			
QY	181	ATTGTTCTTAGTTTCACTACTATTATTATTTCATATTTCATGTTATTTCATGATCGCGCAACGA	240			

Db 754 ATTGTTCTTAGTTTCACTACTATTATTATTTCATATTTCATGTTATTGACATCGCGCAAGCA 813
Qy 241 CCAGCCTCCATACCGATTAGACAGATCTCAACGCTGGGCTCCAGAGCTCACATTTATG 300
Db 814 CCAGCCTCCATACCGATTAGACAGATCTCAACGCTGGGCTCCAGAGCTCACATTTATG 873
Qy 301 CTAATAAATCACTACTACTGTAACAGCTACAGAAAAAATAAAGAGCGAGGATAAA 360
Db 874 CTAATAAATCACTACTACTGTAACAGCTACAGAAAAAATAAAGAGCGAGGATAAA 933
Qy 361 CCACCTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAACCCTTCTTTCTCTCAAA 420
Db 934 CCACCTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAACCCTTCTTTCTCTCAAA 993
Qy 421 ATATCAATAACAGTACTAGTATCAACAGCATATCGAATAAATCACTACTACACAGTAG 480
Db 994 ATATCAATAACAGTACTAGTATCAACAGCATATCGAATAAATCACTACTACACAGTAG 1053
Qy 481 GAACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAACCAACCAACAGGA 540
Db 1054 GAACAGTAACGACAAACGACAGATAGTAACGACAATAACGACAACCAACCAACAGGA 1113
Qy 541 ACACAGATTAAGCTTCAGAAAC 561
Db 1114 ACACAGATTAAGCTTCAGAAAC 1134

RESULT 2
KMAJ76 2198 bp DNA linear PLN 15-APR-2005
LOCUS KMAJ76
DEFINITION Kluyveromyces marxianus epgl gene encoding endopolygalacturonase,
Partial.
ACCESSION AJ000076
VERSION AJ000076.1 GI:2597956
KEYWORDS endopolygalacturonase; epgl gene.
SOURCE Kluyveromyces marxianus
ORGANISM Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1
AUTHORS Siektele, R., Bartkeviute, D. and Sasnauskas, K.
TITLE Cloning, targeted disruption and heterologous expression of the
Kluyveromyces marxianus endopolygalacturonase gene (EPGL)
JOURNAL Yeast 15 (4), 311-322 (1999)
PUBMED 10206190
REFERENCE 2 (bases 1 to 2198)
AUTHORS Siektele, R.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1997) Siektele R., Gene engineering laboratory,
Institute of Biotechnology, V.Graicuno 8, 2028, LITHUANIAN
REPUBLIC
REMARK revised by author 04-NOV-97
FEATURES
source
1..2198
Location/Qualifiers
/organism="Kluyveromyces marxianus"
/mol_type="genomic DNA"
/strain="BMK Y-719"
/db_xref="taxon:4911"
/chromosome="VI"
338..>1655
/gene="epgl"
338..352
/gene="epgl"
573..>1655
/gene="epgl"
/EC_number="3.2.1.15"
/codon_start=1
/product="endopolygalacturonase"
/protein_id="CAA03900.1"
/db_xref="GI:2597957"
/db_xref="GOA:O13478"
/db_xref="InterPro:IPR000743"

gene
TATA_signal
CDS

/db_xref="InterPro:IPR006626"
/db_xref="UniProt/TrEMBL:O13478"
/translation="MLFSNTLLIAAASALLAEASPLEKRDSCSLSGKTAGGGLSNCAT
VTNNVEVPSGKTLDTGLPDGATVNFVQVTFPGDEWVGPLVLSIGKILKVVYKSGH
LIDGCGARWDDGKGGKPKFMSLURLTNSDVGGLQIKNTPIQAISSVNSCSDTVIH
DVTIDNRDGDNLGHTDGFVGNVNTIENSHVYNQDDCIANVSGTGVYFPKNYC
SGHKGASIGVGLRNNVVDVTVYFENNQIVNSDNLRIKTIQKATGSVNNVHFLSNTI
SGIRKFGIIVETDYSNGSTTGTGSKVPIINFEVDGLTGSVDSSAYRVKIFVAGASKW
TKNVDVITGSSSFGSCTGIPSGGAFPC"
573..657
/gene="epgl"
658..>1655
/gene="epgl"
/product="endopolygalacturonase"
/EC_number="3.2.1.15"

ORIGIN

Query Match 99.8%; Score 560; DB 15; Length 2198;
Best Local Similarity 100.0%; Pred. No. 7.4e-75;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 GCAGCCGTTTGGTTACAGTTCTCATGACCAATCAGCCAGATTTTGCATAGTATTAACT 61
Db 1 GCAGCCGTTTGGTTACAGTTCTCATGACCAATCAGCCAGATTTTGCATAGTATTAACT 60
Qy 62 TAGAATTAAGGCAACATCTTTGGATATGATGTAGAGTAAGTCGTTTCGAAACCATATTATTA 121
Db 61 TAGAATTAAGGCAACATCTTTGGATATGATGTAGAGTAAGTCGTTTCGAAACCATATTATTA 120
Qy 122 TTATT 181
Db 121 TTATT 180
Qy 182 TTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 241
Db 181 TTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 240
Qy 242 CAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTCACACATTATGC 301
Db 241 CAGCCTCCATACCGATTAGACAGGATCTCAACGCTGGGCTCCAGAGCTCACACATTATGC 300
Qy 302 TAAATAACTATCTACTGTAACTACAGCTACAGAAAAAATAAAGAGCGAGGATAAAC 361
Db 301 TAAATAACTATCTACTGTAACTACAGCTACAGAAAAAATAAAGAGCGAGGATAAAC 360
Qy 362 CACTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAACCCTTCTTTCTCTCAAAA 421
Db 361 CACTCTCTTGTGAATCAGGATCAGTAGGTAACCTATAAACCCTTCTTTCTCTCAAAA 420
Qy 422 TATCAATAACAGTACTATCAACAGATATCGAATAAATCACTACTACACAGTAGG 481
Db 421 TATCAATAACAGTACTATCAACAGATATCGAATAAATCACTACTACACAGTAGG 480
Qy 482 AACAGTAACGACAAACGACAGATAGTAACGACAATAAAGAGCGAGGATAAAC 541
Db 481 AACAGTAACGACAAACGACAGATAGTAACGACAATAAAGAGCGAGGATAAAC 540
Qy 542 CACAGATTAAGCTCAGAAAC 561
Db 541 CACAGATTAAGCTCAGAAAC 560

RESULT 3
AC116957.1
WPCOMMENT
Sequence split into 4 fragments LOCUS AC116957 Accession AC116957
Fragment Name Begin End
AC116957_0 1 110000
AC116957_1 100001 210000
AC116957_2 200001 310000
AC116957_3 300001 405682
Continuation (2 of 4) of AC116957 from base 100001 (AC116957 Dictyostelium discoideum c.


```
/db_xref="GI:28828451"
/translation="SONQSEDOKLLFGLMYSLKAFITSSPPKKIDDKTGPHCYKTSTY
KLHYETLSCKIFILIMSDPNVPLRDLRLKKIYSQIFVEYIKNPIYKHGTGTYKCDIFI
NQLNLVILKQMPSESS"
CDS
complement(join(3621..3656,3814..3876,3996..4504,
4659..4986))
/notes="GeneID exon scores (in order of location ranges):
0.03, 5.79, 47.43, 25.62 - GSCJ_ID dd_00208"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51035.1"
/db_xref="GI:28828408"
/translation="MKFISYIFIIQIIFINLINGQIVRGPDQLQGVYENGTMILGS
PNVNIIGDQFNMTKOLSNSSWTHRYNVDWDFIRPHRNNQSGFEATDFHPRVDE
FSTIKFSEVGTVMVAEIPYSKNVTSFTSLVDVPCQETLTNGSVVUVDKACONVTD
ITATYQKVENVYSTVLEPILLPKHIFPESRTRFVSFYTIIFQLDNGALVDAWDGD
RFSYCNKDCSDCNQCATKFDMDRDEGVGCOLKFFLAWAGVDKSGKNCVSLNKIPS
KFOYSATAIKNNGTGLVSDFFYKINDNNNPNTA"
CDS
5802..5957
/notes="GeneID exon scores (in order of location ranges):
4.32 - GSCJ_ID dd_00209"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51036.1"
/db_xref="GI:28828409"
/translation="MRKVSYPENHKQSNLENCLTFYQMIQTDPDRSIFKLATSNWL
ISNIKQ"
CDS
complement(join(6544..21168,21250..21832,22134..22616,
22677..23743))
/notes="GeneID exon scores (in order of location ranges):
1211.90, 27.13, 26.28, 60.75 - GSCJ_ID dd_01256"
/pseudo
/codon_start=1
join(24858..25024,25139..25410,25452..25831)
/notes="GeneID exon scores (in order of location ranges):
16.28, 24.00, 42.08 - GSCJ_ID dd_00213"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96709.2"
/db_xref="GI:28828410"
/translation="MSTQGLRFLSLAERKAGSMSTPGAGNSVSNLENDFDNDVKS VH
KTDSTNMLLVGSDSPKILLSTGKINELIEHFDSTSMYAILRLVRVDGIPIT
DRFVYITGLDQVNAIKAKYSVNTKVTLLGHVNVETIASKIEQVTSDDPNMM
LIGYDEKRLGVGTGKGLQSLVHLKQDNVYGVATVDVIGIVOPKHAQITFVG
VSVSPWKGRLTHRGKIDEPFSPHVSFLPSNPQDLTDEILAQVQLKGKMK"
CDS
complement(join(26042..27140,27220..27341,27450..27770,
27854..28531,28636..28935))
/notes="GeneID exon scores (in order of location ranges):
109.13, 7.33, 29.97, 57.67, 11.17 - GSCJ_ID dd_00215"
/codon_start=1
/product="similar to Arabidopsis thaliana (Mouse-ear
cress). Hypothetical 80.1 kDa protein"
/protein_id="AAL96711.2"
/db_xref="GI:28828411"
/translation="MTTNTIINNGVSFSLNNSNIPTIHSPPSNIIVDCIRGTWAY
IYLVDSEYIQOHPNPKPPSLPLPKLYSTTNKTPYPPQFOQNNENISIVKEIY
NYINPEIKDCDFCDVDSIKYARPHNLTLKMGCKSIHARQISNTVFEQLEKCR I
ENQKUTANNAIIVTDSITKQSTTTTTTTTTTTTITATTTQPPYICVSIORN
IFYIIGHLSYCYQSKPOYIIDFPVSCVQDKGHSFTLLGGTQCGKSTIALLAS
RIGFAVISTDNIROLLRKFISQESPIILWASTHAGEIISNPSLHKEKILGYEAO
NMFNKLGDILGHYKREKESLIVGVDLDTKLILRLVKKHPSCIPFLMYISNEAKH
BFAIRSKYMTLDPHONKYTKYFNIRIINDHLCHGADHMIPOIDNTSIDRSLATIH
GTIFACLKRVQCGSRYNHEIDKNNMLNXYEQIQHPWSSKGLRLIQQKTSVPH
GNNVYTGVDNNNSNNNNVNDNGDNGSDNNNDNNNNNNNNNNNNNNNNNNNN
DNNNDNNDNNNDNNDNNDNNSNNSINSNNGNNGNNCVNVNNGNNSDILKVE
YNNKNSNDNENYNSDNDHNIKNETNSDNGINCINNIIISYNNNIKNNTDNTGN
HNYTNTKNDNNDCKNNNNNNIKTIPNEGQOQPOQPOQPOQPOQPOQPOQOQKQO
IQEQQNLNNNNKSIDDEEAFNSDDEHDEDDSIrgnesvgvghsgsvksSLNDK
NHNNGNOSDNDNDDSDISDSDGGERIDYVLGCSLGS"
join(29669..30019,30109..30626,30737..30996,31116..31276,
31353..32132)
/notes="GeneID exon scores (in order of location ranges):
29.16, 56.32, 27.47, 16.22, 92.26 - GSCJ_ID dd_00200"
/
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96712.2"
/db_xref="GI:28828412"
/translation="MSVPSSVIEHNLVPEIPIVYRLNARESNNLITEKEQLYAHHSV
ACWGSKICLGQTSIESGPIFNLFQNLKSTVNPVNISESEYSLLSYAATFY
GMNGLYSGDSKPIPRISKEKQLIINKNDKNVNEYMGKSELMSLDKQVRELGI
DNGLSTVSPNITKVEIKQKFMDSKISIPVNLFPKVSNNYLLASASTSTPT
VSHQDGYTINIYGDNNKULTKVDNLKALPAAENQTNMLKXIIISFYSGSIDD
HKSQRWMIKDISPAVENTIGIESIROPYVGRGEWGFVSVNKEMSUKFGKLTNA
TTFSLKPMWDSFEKENKPDFTSELTFTATTGIPAGINLNDDIRQTEGFKNVSLG
NVIAAGBYETVTFIQESQKLFNELSTAFELQVGHLYHGSGKLFITDANGNVF
KVGAEINPLTNKPIDPKTEVYKFGTYSVFKSLGSPMECKACCGIYLPDPKILE
LFGFTDPKAEADVIVNMLIMARAGVCALEFYPSPSEGAPGKWRQAHMOARCYLITF
LBSGVLTDKADDDIVKLKDSKIRGIVKAVGDFELRLMYKATANIADASKLFBY
THVNEFPLRIDIVLAKPRKRVQAHYILNSNGKVCLOQDFDQKIDSMTIRFG
KDDSDML"
CDS
complement(join(32452..32600,32953..33619))
/notes="GeneID exon scores (in order of location ranges):
-0.20, 32.06 - GSCJ_ID dd_00202"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL96713.2"
/db_xref="GI:28828413"
/translation="MTEVYIQDQIKQFNENPIQKTNKYQYNSDNNSSIVYLK
NPKQNNYNNKICNYKRSLSSSSQFQNEKETRKNKAQFVENDFENGFGNISML
NNSIPKCEDCNGNEPHOKLSODIPLFPPEPLTHEINKVHIELETAGVDKD
DVVDLTNNLITIVAKKSVPLFQNMCEPKHKEKISGVYKRVLFNSNTVOKDTKA
RYVNDSPKAYMACLMYGVDFHINTGLDYCPGPELSQKLIQNDIYQORYOI"
complement(join(34387..34623,34721..34883,34986..35770,
35912..36055,36141..37055,37130..37763,37866..37951))
/notes="GeneID exon scores (in order of location ranges):
13.33, 11.81, 81.94, 11.24, 89.27, 61.07, 13.05 - GSCJ_ID
dd_00204"
/codon_start=1
/product="similar to Mus musculus (Mouse). Adult male
liver tumor cDNA, RIKEN full-length enriched library,
clone:C730036A03 product:ESOPHAGEAL CANCER ASSOCIATED
PROTEIN homolog"
/protein_id="AAO51037.1"
/db_xref="GI:28828414"
/translation="MAERQSASSPTSSPPPOQOQOQTPQPPQYLLNSKNFKVKVNLG
REVERHNPBKKHDPLESETATMGGLAIKENHLNTYVDENFMPTDVLKPSILOQYT
TDIPLNPLBKKHDPLESETATMGGLAIKENHLNTYVDENFMPTDVLKPSILOQYT
SDDNPQVSPMSTGTSKIKIPINRLNKLLEELQDKEDSKSTQSPQDIIIMDLTL
HSELKAAEERVSLKIAIQAKLSLTSIKFSPKFTVATILDTFFGNLVYDRI
KRLQSSKSKHEILLKEQAKETRNWFIASIRELLPRLFVEISILKCYEFGQD
VNTPEQVINRISIRGINPLVANYIRAYLTRSFQDCEPKFVIOQLKDFVFTQ
KSYKSKYLENTLSMYRITLTDYMLGYSPLSLWLOCLAHKATPETLEVLFRSK
NSLLNHIISPPPEYICSNSTMFNFIKADDTLSYKQVLYSTFGVNLVLGQPKNQ
ILSLNDVWVYTPENIKDYISVAEYIEVULHCSEKETDVLKDLIRHLIPPKGY
ETIHSQSVIKFTIHISDFGLVSNFTPLDLDFNGESQKLSRSTLEALTSKY
ETSDPILNTFLTYKALHLSNLSFQDEVQVQLVNVNCINKIDFGDRVSKLNFY
VECTQTFNFDGVKNRLYGVCEICEKTLNLVKGHTPKTTSFIRACVAYCFITPSI
DIFPLNLYLVSSVLAQNALSOADALLKAAITFIOEIPPLEFVKQSTEDMTSWA
YVDFISLLVYTPGHPSEGFYVLYKVIKQWESSSTAKSKFLQILLCLCSWA
QTSUPHIEKVESNDQLPTEPNTLETFFNSLKIILDLNLLKDEPNLTQKV
GLICIDLNLNVLNGLNSLTASLFDNINNAKKIIPSTCFNEITLKNLAFIGILE
SKMQQDIFNKLSSQOQ"
complement(join(38390..42839,42945..43705))
/notes="GeneID exon scores (in order of location ranges):
443.99, 41.08 - GSCJ_ID dd_00207"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Histidine kinase Dhkr"
/protein_id="AAO51038.1"
/db_xref="GI:28828415"
/translation="WVLYKLEKLNQALWDFINKRVLWGNOSCKYKNIESITSLF
NNSLINDIRNCKSGHEIDIESNENGEILKOMIFKYSSIHISMDARQLYLDTSIN
LLVNDQSHLYDEGRKIFILVEAIEIKFQSPSSSENHNINNNQNSVNNSS
LNKQYNPELPSMGSWENWVNDQNTTKASNQFYKIENDFNKNLNFNDINKUGIQE
16.7%; Score 93.6; DB 2; Length 132254;
```

Query Match

Thu Feb 23 13:22:14 2006

[illegible]

Search completed: February 23, 2006, 04:24:42
Job time : 5052.43 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:20:52 ; Search time 492.114 Seconds
(without alignments)
7597.608 Million cell updates/sec

Title: US-10-070-574-1_COPY_574_1134

Perfect score: 561

Sequence: 1 acgacgcgttttgccttag.....cacagattaagctcagaac 561

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3312346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	561	100.0	1146	5	Aaf61989 K. marxia
2	87.4	15.6	2270	4	Abi24848 Drosophil
3	87.4	15.6	5454	3	Aaa70236 Plasmodiu
4	87.2	15.5	4985	6	Abq75107 Anopheles
5	87.2	15.5	4985	10	Acf79720 Mosquito
6	85	15.2	183	12	Adk61708 Base cont
7	85	15.2	1507	13	Adx50603 Plant ful
8	84	15.0	6642	4	Abi05270 Drosophil
9	84	15.0	10992	4	Abi05242 Drosophil
10	84	15.0	11143	4	Abi12834 Drosophil
11	82.2	14.7	529	10	Abz24320 Sclerotin
12	79.4	14.2	2061	3	Aaa70228 Plasmodiu
13	77.6	13.8	1962	14	Adz04202 Plasmodiu
14	77.2	13.8	357	3	Aaz94801 Soybean m
15	77	13.7	412	3	Abn81085 Shrimp po
16	76	13.5	382	3	Abn81050 Shrimp po
17	76	13.5	8413	6	Abi34497 Human met
18	76	13.5	8413	6	Abi70520 Chemical
19	76	13.5	8413	7	AdS99758 Complemen

ALIGNMENTS

RESULT 1

Aaf61989

ID Aaf61989 standard; DNA; 1146 BP.

AC Aaf61989;

DT 22-AUG-2001 (first entry)

DE K. marxianus endopolygalacturonase regulatory region DNA SEQ ID 1.

Endopolygalacturonase; promoter; hepatitis B surface antigen; pectin;

polyoma virus VP1; staphylococcal protein A; vaccine; antiviral;

antibacterial; regulatory region; ds.

OS Kluyveromyces marxianus.

PN WO200120005-A1.

PD 22-MAR-2001.

PF 05-SEP-2000; 2000WO-EP008662.

PR 10-SEP-1999; 99DE-01043383.

XX (TADP-) TAD PHARMA GMBH.

XX Becher D, Sieketele R, Bartkeviciute D, Sasnauskas K, Doehner L;

XX Salim S;

XX WPI; 2001-244812/25.

XX New promoter from Kluyveromyces marxianus, useful for controlling

expression of foreign genes in yeast, e.g. production of vaccine

antigens.

XX Claim 1; Page 26; 32pp; German.

XX This invention describes a novel promoter sequence (S1), isolated from

Kluyveromyces marxianus. (S1) is a promoter for controlling expression of

foreign proteins in yeast, particularly hepatitis B surface antigen,

polyoma virus VP1 or staphylococcal protein A for use in vaccines. (S1)

is a strong promoter that can be induced by adding pectin to the culture

[illegible]

[illegible]


```

ABN81085/c
ID ABN81085 standard; DNA; 412 BP.
XX
AC ABN81085;
XX
DT 16-JUL-2002 (first entry)
XX
DE Shrimp polynucleotide SEQ ID NO 89.
XX
XX Giant black tiger prawn; Penaeus monodon; pacific white shrimp;
KW Litopenaeus vannamei; shrimp; microsatellite sequence; genome mapping;
KW Taura Syndrome Virus; TSV; infection; ds.
XX
OS Penaeus monodon.
XX
PN WO200034476-A2.
XX
PD 15-JUN-2000.
XX
PF 10-DEC-1999; 99WO-US029571.
XX
PR 10-DEC-1998; 98US-0111670P.
XX
PA (TUFT ) TUFTS COLLEGE.
XX
PI Alcivar-Warren A, Xu Z, Dhar AK, Fan Y, Meehan D, Garcia DK;
XX
DR WPI; 2000-423422/36.
XX
PT Polynucleotides of shrimp are useful for identifying, mapping and
PT characterizing of the genome of various species of shrimp.
XX
PS Claim 1; Page 98; 120pp; English.
XX
CC The invention relates to an isolated polynucleotide (I) of the giant
CC black tiger prawn, Penaeus monodon or expressed sequence tags of the
CC pacific white shrimp, Litopenaeus vannamei (ABN80997-ABN81172), both
CC containing microsatellites sequences including those P. monodon
CC microsatellite sequences given in GenBank AF077550-AF077598. (I), the
CC complementary sequence or fragment and the encoded polypeptide are useful
CC for mapping of the genome of various species of shrimp. Mapping the
CC genome of Penaeus is useful for determining whether a test shrimp,
CC preferably Litopenaeus vannamei, has a genotype associated with a
CC phenotypic trait such as resistance to Taura Syndrome Virus (TSV)
CC infection
XX
SQ Sequence 412 BP; 213 A; 30 C; 41 G; 128 T; 0 U; 0 Other;
Query Match 13.7%; Score 77; DB 3; Length 412;
Best Local Similarity 76.0%; Pred No. 0.00028;
Matches 95; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Qy 85 ATATGATGTAGAGTAAGTCGTCGAAACCACTATTATTATTATTATTATTATTATTATT 144
Db 192 ATTATCATTAATAAATACCATCGTTATTATTATTATTATTATTATTATTATTATT 133
Qy 145 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 204
Db 132 ATTATTGTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 73
Qy 205 ATTAT 209
Db 72 ATTAT 68

```

Search completed: February 23, 2006, 01:52:07
Job time : 494.114 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2006, 01:40:58 ; Search time 3385.83 Seconds
(without alignments)
7752.175 Million cell updates/sec

Title: US-10-070-574-1_COPY_574_1134

Perfect score: 561

Sequence: 1 agcagcgcttttcttacag.....cacagattaagctcagaac 561

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.8	16.7	702	10	CW161301
2	93.4	16.6	891	10	CW952279
3	92.2	16.4	579	1	AU038989
4	92.2	16.4	655	4	AY068594
5	92.2	16.4	744	10	CW161302
6	92	16.4	580	9	BZ639429
7	91.8	16.4	435	10	CW039478
8	91.6	16.3	832	9	BH391984
9	91.2	16.3	762	10	CL992030
10	91.2	16.3	876	9	BZ700706
11	91.2	16.3	985	10	CG205910
12	90.8	16.2	701	10	CG200505
13	90.8	16.2	946	10	CG814583
14	90.6	16.1	713	9	AZ217550
15	90.6	16.1	891	10	CW944789
16	90.4	16.1	637	10	CW293181
17	90.4	16.1	674	10	CL705363
18	90.4	16.1	679	10	CW233591
19	90.4	16.1	843	4	AY068764
20	90.2	16.1	617	3	BJ394972
21	90	16.0	693	10	CW189159
22	89.8	16.0	622	10	CW291169

C	23	89.6	16.0	445	3	BJ392338
	24	89.6	16.0	474	10	CG106547
C	25	89.6	16.0	590	3	BJ365627
	26	89.4	15.9	480	9	BZ393762
C	27	89.4	15.9	624	10	CL801804
C	28	89.4	15.9	711	9	BH969990
C	29	89.2	15.9	311	1	AU054171
C	30	89.2	15.9	546	9	BZ999891
	31	89.2	15.9	839	10	BX168691
	32	88.8	15.8	520	8	DN550639
	33	88.8	15.8	541	8	DN550960
	34	88.8	15.8	565	8	DN203124
	35	88.8	15.8	617	8	DN551066
	36	88.8	15.8	673	10	CW024383
	37	88.8	15.8	690	10	CW165754
	38	88.6	15.8	305	1	AU038805
	39	88.6	15.8	750	10	CG033026
	40	88.6	15.8	832	10	CG136380
	41	88.6	15.8	989	11	CNS032JH
	42	88.4	15.8	352	10	CW965330
	43	88.4	15.8	654	10	CW372547
	44	88.4	15.8	687	9	BH401326
	45	88.4	15.8	738	10	CL159291

ALIGNMENTS

RESULT 1
CW161301
LOCUS
DEFINITION
104_568_11150342_116_36436_061 Sorghum bicolor genomic clone 11150342, genomic survey sequence.
ACCESSION
CW161301
VERSION
GI:54853848
KEYWORDS
Sorghum bicolor (sorghum)
SOURCE
Sorghum bicolor
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 702)
Bedell, J.A., Budiman, M.A., Nunberg, A., Citek, R.W., Robbins, D., Jones, J., Flick, E., Rohlfing, T., Fries, J., Bradford, K., McMenamy, J., Smith, M., Holsman, H., Roe, B.A., Wiley, G., Korf, I.F., Rabinowicz, P.D., Lakey, N., McCombie, W.R., Jeddloch, J.A. and Martienssen, R.A.
Sorghum genome sequencing by methylation filtration
PLOS Biol. 3 (1), e13 (2005)
15660154
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: 568 row: d column: 14
Seq primer: T3 Reverse
Class: methylation filtered
High quality sequence stop: 702.
Location/Qualifiers
1..702
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="MTx623"
/db_xref="taxon:4558"
/clone="11150342"
/clone_lib="Sorghum methylation filtered library (LibID: 104)"
/note="Organ: leaf; Vector: pBCSK(-); Site 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5

Class: methylation filtered
High quality sequence stop: 744.
Location/Qualifiers
1. .744
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/cultivar="ATx623"
/db_xref="taxon:4558"
/clone="11150342"
/clone_lib="Sorghum methylation filtered library (LibID: 104)"
/note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pBCSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN
Query Match 16.4%; Score 92.2; DB 10; Length 744;
Best Local Similarity 70.1%; Pred. No. 4.9e-09;
Matches 124; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 54 TATTAACCTTAGAATTAAGCACAATCTTGGATATGCATGTAGATGAAGTCGTTCCAAAC 113
|||||
Db 614 TATCAAAAAAAGTAAGTAAACAGCACCACCAATTTCAAGTTGTTTATTATTATTATTAT 555
|||||
QY 114 CATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 173
|||||
Db 554 TAT 495
|||||
QY 174 AATTGTTATTGTTCTTAGTTTCACACTATTATTATTATTCATATTCATGTTATTGCACAT 230
|||||
Db 494 TAT 438
|||||

RESULT 6
BZ639429/c
LOCUS
DEFINITION BZ639429 580 bp DNA linear GSS 29-JAN-2003
OGANO65TM_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBma0102K09,
genomic survey sequence.
ACCESSION BZ639429
VERSION BZ639429.1 GI:28093531
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 580)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reenick,A., Frazer,C.M., Budinan,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
COMMENT Other GSSs: OGANO65TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: methylation filtered.
Location/Qualifiers
1. .580
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBma0102K09"
/clone_lib="ZM_0.7_1.5_KB"
/note="Vector: pBCSK(-); Site 1: HincII; 0.7-1.5 kb

FEATURES
source

source

1. .946
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Forrest"
/db_xref="taxon:3847"
/clones="B10D13:MTPLB6"
/clone_lib="LargeInsertSoybeanGenLib"
/notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 EcoRI clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.) cv. Forrest seeds were grown in greenhouse for fourteen days. Nuclei were isolated and embedded in agarose, restriction digested with Hind III BamHI or EcoRI, large size DNA fragments were ligated in vector V41 (pCLD04541) and electro transformed in DH10a cells. About 90,000 clones from BAC libraries were fingerprinted with HindIII and Hae III. Version 2 (automatic build) Contigs were built from 78,001 fingerprints. Contigs were manually examined to find the best non redundant tile path through the contigs representing about 13,000 clones. The clones were end sequenced."

ORIGIN

Query Match 16.2%; Score 90.8; DB 10; Length 946;
Best Local Similarity 70.1%; Pred. No. 9.6e-09;
Matches 122; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 52 AGTATTAACTAGCAATTAAGCCACATCTTTGGATGATGATGAGTAAGTCGTTGCA 111
|||||
Db 771 ATTATTATTAATAAATTTCCATCATGTTTATTACTACTAATTTCTAGCGTTGTTGGT 830
|||||
QY 112 ACCATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 171
|||||
Db 831 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 890
|||||
QY 172 GAAATTGTTATGTTCTTCTAGTTTCACTATTTATTATTATTCATATTCATGTTATT 225
|||||
Db 891 ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 944
|||||

RESULT 14

AZ217550/c
LOCUS
DEFINITION
Sheared DNA-86A9.TF Sheared DNA Trypanosoma brucei genomic clone

ACCESSION
AZ217550
VERSION
GSS.
KEYWORDS
SOURCE
Trypanosoma brucei
Trypanosoma brucei
Trypanosoma
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

ORGANISM
Trypanosoma
1 (bases 1 to 713)
REFERENCE
AUTHORS
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
Donelson,J., Fraser,C. and Adams,M.

TITLE
Determination of clone end sequences from Trypanosoma brucei GUTat
10-1 sheared DNA library
JOURNAL
Unpublished (1999)
COMMENT
Other_GSSs: Sheared DNA-86A9.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10-1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.

Location/Qualifiers
1. .891

ORGANISM="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"

FEATURES
source

1. .891
Location/Qualifiers
1. .891

ORGANISM="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"

source

1. .713
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-86A9"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

ORIGIN

Query Match 16.1%; Score 90.6; DB 9; Length 713;
Best Local Similarity 62.7%; Pred. No. 1.1e-08;
Matches 141; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 6 CGTTTTGTTACAGTCTCTATGCAATTCAGCCAGATTTTGCATAGTATTAACTAGA 65
|||||
Db 324 CCAATTTCTTCAAAATTTATTATATAGGCCTCAAAATACAAATCTTTATTATTAT 265
|||||
QY 66 APTAAGGCAACATCTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 125
|||||
Db 264 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 205
|||||
QY 126 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 185
|||||
Db 204 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 145
|||||
QY 186 TCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTATTATT 230
|||||
Db 144 TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 100
|||||

RESULT 15

LOCUS
CW944789/c
DEFINITION
TC827.1_B01.T7 Tribolium BAC library Tribolium castaneum genomic,
genomic survey sequence.

ACCESSION
CW944789
VERSION
GSS.
KEYWORDS
SOURCE
Tribolium castaneum (red flour beetle)

ORGANISM
Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.

REFERENCE
1 (bases 1 to 891)
AUTHORS
Savard J. and Tautz, D.
TITLE
Tribolium castaneum BAC-ends sequencing project
JOURNAL
Unpublished (2003)
COMMENT
Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universität zu Köln
Weyertal 121, 50931 Köln, Germany
Tel: 49 221 470 5971
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.

Location/Qualifiers
1. .891

ORGANISM="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"

FEATURES
source

1. .891
Location/Qualifiers
1. .891

ORGANISM="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"

FEATURES
source

/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN

```
Query Match      16.1%; Score 90.6; DB 10; Length 891;  
Best Local Similarity 71.0%; Pred. No. 1.1e-08;  
Matches 120; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
  
Qy 62 TAGAATTAAAGGCAACATCTTTGGATATGCATGTAGAGTAAGTCGTTGGAACCATTA 121  
   |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  
Db 395 TAAAAATTTCCAAAGCCTGTTTCAGTTTGAAATTCCTGTATTTCATTTATTATTATTA 336  
  
Qy 122 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 181  
   |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  
Db 335 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTA 276  
  
Qy 182 TTGTTCTTAGTTTCACTACTATTATTATTATTATTATTATTATTATTATTATTGACAT 230  
   |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  
Db 275 TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 227
```

Search completed: February 23, 2006, 03:40:23
Job time : 3389.83 secs